User Guide for ELSA Geographical Variables

This User Guide details the geographical variables that are available to be matched to ELSA data, as well as detailing the sources of these variables. This guide is not exhaustive – there are many other geographical variables available but here we have detailed those that we have already matched on to the ELSA analytical identifier (IDAUNIQ). These variables comprise the 'ELSA geography file'. Many of these variables can be considered to be 'disclosive', hence why they are not archived with the rest of the ELSA data and are only available by special request. The first section of this guide describes the basic details, the second section gives detailed information about the variables in the ELSA geography file and the third section gives information about how to obtain these variables.

SECTION A Basic Information

The geography file contains geographically related information in the form of identifiers such as Government Office Region (GOR), and indices that are geographically based such as the Index of Multiple Deprivation (IMD). An ELSA respondent's geography is determined by the household that they reside in, more specifically their postcode. The nature of ELSA implies that geography can change over time so that the information in this file is also time dependent. Key points in time have been selected (i.e. survey interview dates), and geographical information is given for those key points. That is, the geographical information attached to a record relates to the interview address. More specifically, we use the postcode of the interview address to match in geographical information such as GOR and IMD.[1]

Note that a respondent can live in other places between interviews but this information is not represented in the file (and is applicable to a relatively small proportion of respondents). Currently there are three key points in time, HSE (Wave 0) and ELSA Waves 1 and 2. Hence there is a separate geography file for each Wave. Note also that the geography file for Wave *t* only contains the records for responding individuals at Wave *t* (n=19,834 at Wave 0, n=12,099 at Wave 1 and n=9,432 at Wave 2).

The analytical identifier supplied for matching to archived ELSA data is IDAUNIQ. Once a respondent has been assigned this identifier its value remains constant throughout the following Waves.

[1] The postcodes of ELSA respondents will **never** be given out to external researchers. Under strict conditions we may be able to supply geographical identifiers such as postcode sector – this may involve researchers having to come and work in the NatCen Data Enclave. Please see Section C for more details.

SECTION B Variable Descriptions

1. Geographical identifiers

The table below sets out the geographical variables already attached to the ELSA geography files (X denotes Wave number).

The final column shows the level of disclosivity, which is defined further as follows:

LOW - If the applicant has justified the need for these variables, we will usually provide them with the ELSA analytical serial number (IDAUNIQ) attached.

HIGH - We will not generally give out this information but instead will try and find an alternative way of enabling the applicant to do their analyses (see section C for details).

Variable	Variable label	Original source	Matching variable	Level of disclosivity
oacode	2001 Census Output Area	National Statstics Postcode Directory (NSPD)	Postcode	(Very) High
wXpsect	Postcode sector (interview address)	Royal Mail Postcode Address File (PAF)	Postcode	High
easting	National Grid reference – Easting	NSPD	Postcode	High
northing	National Grid reference – Northing	NSPD	Postcode	High
lafrCurr	Local Authority	Derived by merging OSCTY & OSLAUA from the National Statistics Postcode Directory *	Postcode	Low
laid	Type of Local Authority: 1 City of London & London Boroughs 2 English Metropolitan districts 3 English non-metropolitan districts 4 English Unitary Authorities 5 Welsh Unitary Authorities	http://www.statistics.gov.uk/geography/downloads/UK_Local_Authorities_2001.xls	LafrCurr	Low
laname	Name of Local Authority	http://www.statistics.gov.uk/geography/downloads/UK_Local_Authorities_2001.xls	LAfrCurr	Low
gor	Government office region A North East B North West D Yorkshire and The Humber E East Midlands F West Midlands G East of England H London J South East K South West	NSPD	Postcode	Low
streg	Standard region 1 North 2 Yorkshire and Humberside Durham 3 East Midlands 4 East Anglia 5 South East 6 South West 7 West Midlands 8 North West 9 Wales	NSPD	Postcode	Low

* OSCTY is County (Office for National Statistics) code, OSLAUA is Local Authority/Unitary Authority.

2. Population density (postcode sector level)

The source was Census data (which NatCen has matched on to the Postcode Address File) and the matching variable was postcode sector.

Variable	Variable label	Level of disclosivity
popdens	Population density of the postcode sector (number of persons in private households per hectare)	High
qpopdens	Population density of the postcode sector (quintiles) 1 least dense 2 2nd quintile 3 3rd quintile 4 4th quintile 5 most dense	Low

Despite being matched successfully using the Wave 0 (HSE) or Wave 1 postcode, a number of cases could not matched to the population density variable using the latest postcode sector file as the lookup due to the postcodes being terminated. Two strategies were used to compute population density for the missing cases:

1. Previous PAF files were used as look-up tables; and

2. For those cases still without a match, population density was imputed as the mean value for those postcodes nested within the same Census Output Area.

The population density quintile was assigned on the basis of the full PAF cut-offs i.e. the whole of England.

3. Department of the Environment, Food and Rural Affairs (DEFRA) urban/rural indicator (Census Output Area level)

The source was the National Statistics Postcode Directory and the matching variable was postcode.

Variable	Variable label	Level of disclosivity
urindew	1 Urban > 10k (sparse) 2 Town & Fringe (sparse) 3 Village (sparse) 4 Hamlet & isolated dwelling (sparse) 5 Urban > 10k (less sparse) 6 Town & Fringe (less sparse) 7 Village (less sparse) 8 Hamlet & isolated dwelling (less sparse)	Low

Note that a four category indicator can easily be derived from this variable by merging categories 1 and 5 (urban >10k), 2 and 6 (Town and Fringe), 3 and 7 (village), and 4 and 8 (Hamlet and Isolated dwelling).

4. Index of Multiple Deprivation (IMD)

A small number of cases (1 in Wave 0 and 8 in Wave 2) have missing values because the address was located within Wales (which has its own Index).

The 2004 Index of Multiple Deprivation

The raw 2004 IMD scores can be found at:

http://www.communities.gov.uk/index.asp?id=1128440

The matching variable was Lower Super Output Area (not provided with the geography data).

Variable	Variable label	Level of disclosivity
imd2004	Index of Multiple Deprivation (2004)	High
qimd2004	Index of Multiple Deprivation 2004 (quintiles) 1 Least deprived 2 2nd quintile 3 3rd quintile 4 4th quintile 5 Most deprived	Low

The 2000 Index of Multiple Deprivation

The raw 2000 IMD scores can be found via a download at:

http://www.communities.gov.uk/index.asp?id=1128626

The matching variable was WARD98 (provided by the NSPD, not provided with the geography data). WARD98 is the 1998 administrative and electoral area.

Variable	Variable label	Level of disclosivity
imd2000	Index of Multiple Deprivation (2000)	High
qimd2000	Index of Multiple Deprivation 2000 (quintiles) 1 Least deprived 2 2nd quintile 3 3rd quintile 4 4th quintile 5 Most deprived	Low

5. Additional variables

It is not possible to anticipate all the various geographically related information that users of the data may require. The geographical variables provided are a subset of those NatCen would be able to supply on request - subject to approval by the NatCen Data Release Panel. A list of some additional variables is provided in the table below:

Variable	Variable label	Original source	Matching variable	Level of disclosivity
casward[1]	Census Area Statistics (CAS) wards	All Fields Postcode Directory (AFPD), now known as National Statistics Postcode Directory (NSPD)	Postcode	High
statward	Statistical ward	AFPD	Postcode	High
oshlthau	Strategic Health Authority	AFPD	Postcode	Low
pct	Primary Care Trust	AFPD	Postcode	Low

[1] for more detail on the exact definition of wards see:

http://www.statistics.gov.uk/geography/statistical_cas_st_wards.asp

SECTION C PROCESS OF OBTAINING GEOGRAPHICAL VARIABLES

If you wish to apply for any of the geographical variables detailed above then please get in touch with the ELSA Data Manager (contact details below) in order to obtain an application form. This application form, once completed and received back at NatCen, (and any initial queries have been resolved) will go to the NatCen Data Release Panel for consideration. On this form, you will need to specify the variables you need, for which Waves of ELSA data you need them, why you need the variables and who exactly will have access to them. You will also need to agree a timeframe in which a) you would like to receive the data (please give as much notice as possible) and b) when you will no longer need the data, and will therefore be obliged to destroy it. You will also need to agree to our standards of security in order to protect the data that you are sent.

The NatCen Data Release Panel meets once a month, and here, if your application is approved, your request will either be agreed as it is or necessary amendments will be suggested. Once this decision has been relayed to you, and you are happy with it, we will need a signed copy of the final form before your data can be sent to you.

Please note that NatCen does **not** give out data that allows low level geographies to be identified i.e. post code sector or ward. If data matched at this level is requested we will attempt to find an alternative way of enabling you to do your analysis without providing data that identifies local areas to you. If you wish to work with particularly disclosive or confidential variables (e.g. postcode sector, grid reference data) then we may suggest that you come and do your analyses in the NatCen Data Enclave. This is a 'safe setting' in which to work with sensitive data, in that the computer in the Enclave is not networked and can only be accessed by a small number of people. Alternatives to using the Enclave are using data with an anonymised identifier attached, so that the ELSA identifier (IDAUNIQ) and the sensitive variable(s) are never seen by the external researcher on the same dataset.

SECTION D CONTACT DETAILS

ELSA Research Team:
E-mail: ELSAData@natcen.ac.uk
Telephone: 020 7250 1866